1-7-2014

A unique covalent bond in basement membrane is a primordial innovation for tissue evolution

Aaron L. Fidler
Vanderbilt University

Roberto M. Vanacore
Vanderbilt University

Sergei V. Chetyrkin
Vanderbilt University

Vadim K. Pedchenko
Vanderbilt University

Gautam Bhave
Vanderbilt University

See next page for additional authors

Recommended Citation

Aaron L. Fidler, Roberto M. Vanacore, Sergei V. Chetyrkin, Vadim K. Pedchenko, Gautam Bhave, Viravuth P. Yin, Cody L. Stothers, Kristie Lindsey Rose, W. Hayes McDonald, Travis A. Clark, Dorin-Bogdan Borza, Robert E. Steele, Michael T. Ivy, The Aspirnauts, Julie K. Hudson, Billy G. Hudson "A triad in the ECM essential for tissue evolution", Proceedings of the National Academy of Sciences Jan 2014, 111 (1) 331-336; DOI: 10.1073/pnas.1318499111

Authors

Aaron L. Fidler, Roberto M. Vanacore, Sergei V. Chetyrkin, Vadim K. Pedchenko, Gautam Bhave, Viravuth P. Yin, Cody L. Stothers, Kristie Lindsey Rose, W. Hayes McDonald, Travis A. Clark, Dorin-Bogdan Borza, Robert E. Steele, Michael T. Ivy, The Aspirnauts, Julie K. Hudson, and Billy G. Hudson

A unique covalent bond in basement membrane is a primordial innovation for tissue evolution

Aaron L. Fidler[a,1], Roberto M. Vanacore[a,b,1], Sergei V. Chetyrkin[a,b,1], Vadim K. Pedchenko[a,b], Gautam Bhave[a,b], Viravuth P. Yin[c], Cody L. Stothers[a], Kristie Lindsey Rose[d,e], W. Hayes McDonald[d,e,f], Travis A. Clark[g], Dorin-Bogdan Borza[a,b], Robert E. Steele[h], Michael T. Ivy[i], The Aspirnauts[b,2], Julie K. Hudson[j], and Billy G. Hudson[a,b,c,d,f,k,l,3]

[a]Department of Medicine, Division of Nephrology and Hypertension, [b]Center for Matrix Biology, [d]Department of Biochemistry, [e]Mass Spectrometry Research Center, [f]Vanderbilt–Ingram Cancer Center, [g]Vanderbilt Technologies for Advanced Genomics, [j]Department of Medical Education and Administration, [k]Department of Pathology, Microbiology, and Immunology, and [l]Vanderbilt Institute of Chemical Biology, Vanderbilt University Medical Center, Nashville, TN 37232; [c]Kathryn W. Davis Center for Regenerative Biology and Medicine, Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672; [h]Department of Biological Chemistry, University of California, Irvine, CA 92697; and [i]Department of Biological Sciences, Tennessee State University, Nashville, TN 37209

Edited* by Mina J. Bissell, E. O. Lawrence Berkeley National Laboratory, Berkeley, CA, and approved November 22, 2013 (received for review September 30, 2013)

Basement membrane, a specialized ECM that underlies polarized epithelium of eumetazoans, provides signaling cues that regulate cell behavior and function in tissue genesis and homeostasis. A collagen IV scaffold, a major component, is essential for tissues and dysfunctional in several diseases. Studies of bovine and *Drosophila* tissues reveal that the scaffold is stabilized by sulfilimine chemical bonds (S = N) that covalently cross-link methionine and hydroxylysine residues at the interface of adjoining triple helical protomers. Peroxidasin, a heme peroxidase embedded in the basement membrane, produces hypohalous acid intermediates that oxidize methionine, forming the sulfilimine cross-link. We explored whether the sulfilimine cross-link is a fundamental requirement in the genesis and evolution of epithelial tissues by determining its occurrence and evolutionary origin in Eumetazoa and its essentiality in zebrafish development; 31 species, spanning 11 major phyla, were investigated for the occurrence of the sulfilimine cross-link by electrophoresis, MS, and multiple sequence alignment of de novo transcriptome and available genomic data for collagen IV and peroxidasin. The results show that the cross-link is conserved throughout Eumetazoa and arose at the divergence of Porifera and Cnidaria over 500 Mya. Also, peroxidasin, the enzyme that forms the bond, is evolutionarily conserved throughout Metazoa. Morpholino knockdown of peroxidasin in zebrafish revealed that the cross-link is essential for organogenesis. Collectively, our findings establish that the triad—a collagen IV scaffold with sulfilimine cross-links, peroxidasin, and hypohalous acids—is a primordial innovation of the ECM essential for organogenesis and tissue evolution.

The extracellular matrix (ECM) provides signaling cues that regulate cell behavior and function in tissue genesis and homeostasis (1–3). A specialized form of ECM, the basement membrane, underlies a layer of polarized cells, forming a basic architectural feature of animal tissues. Basement membranes serve as scaffolds for cell migration and adhesion, delineate apical–basal polarity, and modulate cell differentiation during development (1–7), and in the form of decellularized scaffolds, they guide pluripotent cells to partially regenerate whole organs (8–12). A major component is a collagen IV scaffold that is essential for tissue genesis and dysfunctional in several diseases (13–16). This scaffold is characterized by a network of oligomerized triple helical molecules (Fig. 1*A*) that confer structural integrity to tissues, serves as a ligand for cell surface receptors, such as integrins to modify cell behavior, and serves as a locus for bone morphogenetic protein gradients for patterning in tissue development (17–19).

We recently discovered that the collagen IV scaffolds of bovine, mouse, and *Drosophila melanogaster* tissues are stabilized by sulfilimine chemical bonds (S = N) (20–22). The bond is a covalent cross-link between methionine-93 (Met^{93}) and lysine-211 or hydroxylysine-211 (Lys^{211}/Hyl^{211}) residues that stabilizes the noncollagenous (NC1) interface of adjoining triple helical protomers (Fig. 1*A*). Peroxidasin, an animal heme peroxidase embedded in basement membrane, forms these sulfilimine cross-links by producing hypohalous acids as intermediates for Met^{93} oxidization (21, 22). Here, we explored whether the sulfilimine cross-link of collagen IV is a fundamental requirement in the genesis and evolution of tissues in Eumetazoa. We show the cnidarian origin and essentiality of the sulfilimine cross-link in tissue development and the cnidarian origin of the paradoxical anabolic function of hypohalous acids in tissue genesis.

Results and Discussion

Evolutionary Origin of the Sulfilimine Bond. We sought to determine the occurrence and origin of the sulfilimine cross-link within

Significance

The evolution of multicellular animals from single-celled ancestors was one of the most significant transitions of life on earth. The emergence of larger, more complex animals able to resist predation and colonize new environments was enabled, in part, by a collagen scaffold, which anchors cells together to form tissues and organs. Here, we show that a unique chemical bond, a link between sulfur and nitrogen atoms called a sulfilimine bond, arose over 500 Mya, binding this scaffold together and enabling tissues to withstand mechanical forces. Peroxidasin forms the bond by generating hypohalous acids as strong oxidants, a form of bleach, which normally function as antimicrobial agents. These understandings may lead to approaches for targeting tumors and treatment of other diseases.

Author contributions: R.M.V., S.V.C., V.K.P., V.P.Y., M.T.I., J.K.H., and B.G.H. designed research; A.L.F., S.V.C., G.B., V.P.Y., C.L.S., K.L.R., W.H.M., T.A.C., D.-B.B., R.E.S., and T.A. performed research; G.B. contributed new reagents/analytic tools; A.L.F., R.M.V., S.V.C., V.K.P., V.P.Y., D.-B.B., and R.E.S. analyzed data; and A.L.F. and B.G.H. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GAMX01000001, GAMX01000002, GAND01000001, GAND01000002, GANB01000001, GANB01000002, GAMY01000001, GAMY01000002, GANA01000001, GANA01000002, GAMZ01000002, and GANC01000002).

[1]A.L.F., R.M.V., and S.V.C. contributed equally to this work.

[2]A list of The Aspirnaut coauthors can be found in Table S2. Aspirnaut is a K–20 Science, Technology, Engineering, and Math (STEM) pipeline program for diversity that partners the experiential and content expertise of Vanderbilt University with rural kindergarten through 12th grade schools and diverse high school, undergraduate, and graduate students.

[3]To whom correspondence should be addressed. E-mail: billy.hudson@vanderbilt.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1318499111/-/DCSupplemental.

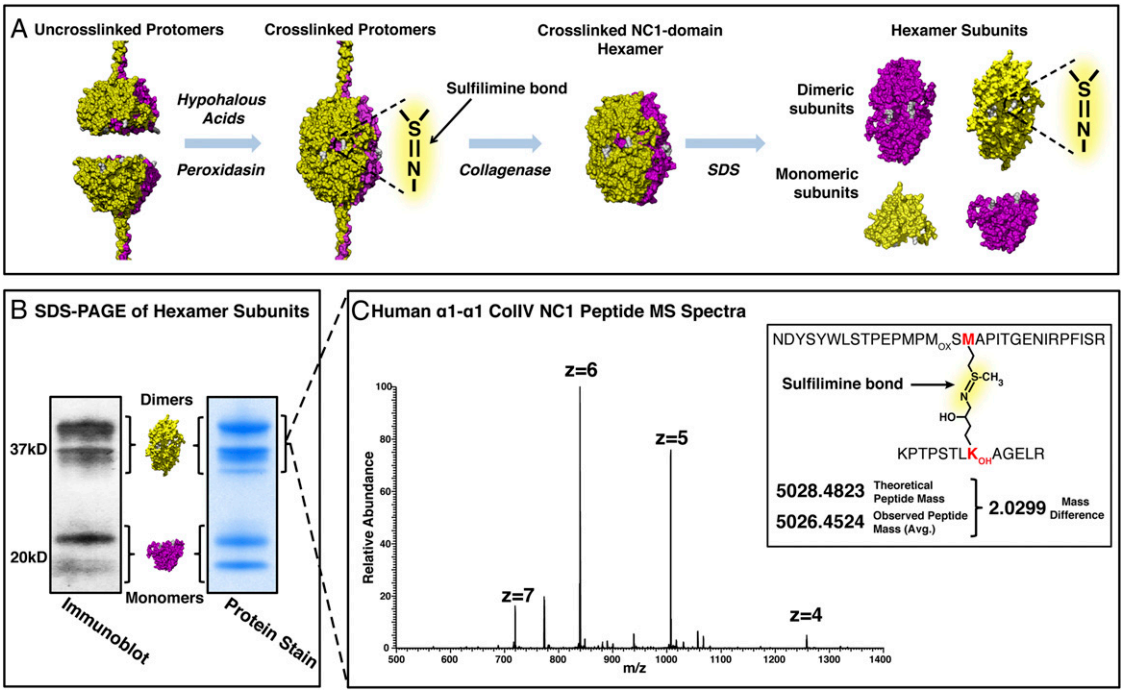

Fig. 1. The sulfilimine bond stabilizes collagen IV scaffolds by the cross-linking of triple helical building block protomers. (*A*) The sulfilimine bond cross-links Met^{93} and Hyl^{211} at the interface between the trimeric NC1 domains of two adjoining protomers, forming a globular hexamer structure. (*B*) Dimeric subunits reflect the presence of the sulfilimine bond in human collagen IV by immunoblot (JK2 Ab) and protein stain. (*C*) MS analysis of tryptic peptides derived from dimeric subunits verified the presence of the bond by a mass difference of 2.0299 between theoretical mass of uncross-linked and observed mass of cross-linked peptides and subsequent multistep CID fragmentation (MS^2/MS^3) analyses.

Metazoa by investigating over 30 species spanning the major animal phyla. Unambiguous detection of the bond required knowledge of NC1 domain primary structure for each species under study, identification of NC1 dimer subunits by SDS/PAGE, and mass spectroscopic analysis of tryptic peptides derived from the cross-linked dimer subunits that putatively contain the sulfilimine cross-link. This approach is presented in Fig. 1 for the human collagen IV scaffold. NC1 hexamers, excised by collagenase digestion, were dissociated into dimeric and monomeric subunits by SDS/PAGE (Fig. 1*A*) and subsequently visualized by protein stain and immunoblot (Fig. 1*B*). The presence of dimer subunits indicates presence of a sulfilimine cross-link. To unambiguously establish the presence of the cross-link in the dimer subunits, tryptic peptides containing Met^{93} and Hyl^{211} were analyzed by high-resolution Orbitrap MS (Orbitrap) as previously described (20). A loss of mass equivalent of two hydrogen atoms was observed between the theoretical mass of an uncross-linked peptide and the observed mass of the cross-linked peptide (Fig. 1*A* and Fig. S1*A*). Furthermore, olefin and methylsulfonamide fragment products generated by multistep collision-induced dissociation (CID) fragmentation (MS^2/MS^3) analyses of multiple charge state ions were observed (Fig. S1 *B–D*). Together, these data provided unequivocal chemical evidence for the presence of the sulfilimine cross-link, which was previously delineated for bovine, murine, and *Drosophila* NC1 domains (20, 21).

Additionally, 30 more species, representing 11 major phyla spanning all of Metazoa, were investigated for this cross-link. Genomic data were unavailable for four of the cnidarian species plus the earthworm, *Lumbricus terrestris*. For these species, next generation RNA sequencing (RNA-Seq) was performed to determine the primary structure of NC1 domains and the presence or absence of Met^{93} and Lys/Hyl^{211}, which was required for Fourier-transform MS (FTMS) analyses of tryptic peptides that encompasses the putative bond. Multiple sequence alignment of animal collagen IV NC1 domains indicated that the requisite Met^{93} and Lys/Hyl^{211} residues were conserved in all eumetazoan phyla, with the exception of the cnidarian *Hydra magnipapillata*, but absent in the noneumetazoan groups Porifera, Placozoa, and Choanozoa (Fig. 2*A*).

Electrophoresis and MS were used to determine the occurrence of the sulfilimine bond/cross-link across the major eumetazoan phyla. Characterization of the NC1 domains by SDS/PAGE revealed the presence of dimer subunits, indicative of sulfilimine cross-links, in all species except for *H. magnipapillata* (Fig. 3 and Fig. S2). These findings are congruent with the conservation of Met^{93} and Hyl^{211} residues across Eumetazoa (Fig. 2*A*). Definitive evidence for the cross-link among species of the major phyla spanning Radiata, Protostoma, and Deuterostoma was achieved by FTMS analysis of trypsin-digested NC1 peptides. The analyses for *Macaca mulatta*, *Danio rerio*, *Ascaris suum*, *Caenorhabditis elegans*, and *Nematostella vectensis* are shown in Figs. S3, S4, S5, S6, and S7, respectively. In each case, loss of two hydrogen atoms between the theoretical and the observed peptide mass and the CID fragmentation pattern of the tryptic peptide containing Met^{93} and Hyl^{211} showed the presence of the sulfilimine cross-link. Collectively, the analyses reveal the occurrence of the sulfilimine cross-link in all of the major eumetazoan phyla, and the appearance of the cross-link coincided with the divergence of Porifera and Cnidaria. Importantly, these findings extend earlier studies of collagen IV in invertebrates (23–26) and reveal that the collagen IV scaffold itself, stabilized by sulfilimine cross-links, is conserved throughout eumetazoan phyla.

Peroxidasin, the enzyme that forms the sulfilimine cross-link, is also evolutionarily conserved throughout Eumetazoa. Our study delineates peroxidasin and its domain structure from Protostoma (27) basally to Radiata and Placozoa. The primary structure for peroxidasin was unknown or incomplete in six cnidarian species, which were determined by RNA-Seq. The multidomain feature of peroxidasin [the peroxidase domain, leucine-rich repeat domains, and Ig-like domains] is conserved within cnidarians and bilaterians alike (Fig. 2*B*). Interestingly,

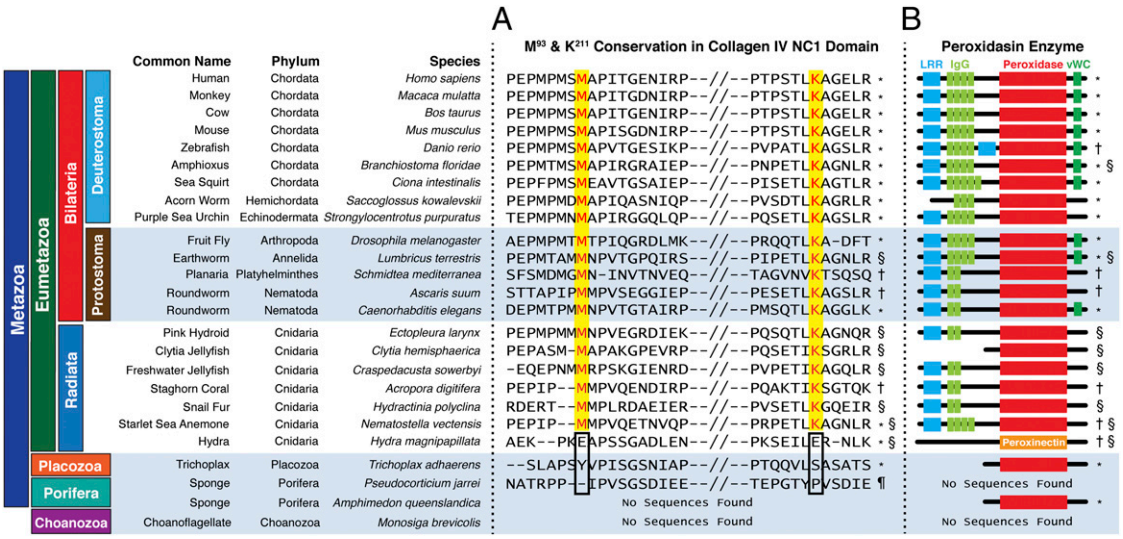

Fig. 2. Multiple sequence alignment of collagen IV NC1 domains encompassing Met^{93} and Hyl^{211} amino acid residues and Pxdn among 11 metazoan and 1 protozoan phyla. (*A*) Met^{93} and Lys/Hyl^{211} (yellow) are conserved in all eumetazoans, except for the cnidarian *H. magnipapillata*, and they are absent in the phyla of Placozoa and Porifera and the protozoan phylum Choanozoa. All sequences belong to the collagen IV α1-like subfamily of chains, except for *Drosophila* (viking) and *Ascaris* (α2 chain). (*B*) Schematic representations of Pxdn. Pxdn sequence was incomplete on both ends for *Mytilus*, *Clytia*, *Trichoplax*, and *Monosiga* and short on one end for *Saccoglossus*, which is indicated here by a shortened schematic representation. Sequence data were gathered from *National Center for Biotechnology Information Reference Sequence, †gathered from whole-genome shotgun/transcriptome shotgun assembly, §generated by RNA-Seq analysis of animal tissues, or ¶assembled from cDNA libraries. All National Center for Biotechnology Information GenBank accession numbers are listed in Table S1.

the von Willebrand factor type C domain is only conserved within Bilateria, suggesting a role in bilaterian innovations such as tubular epithelia, circulatory systems, or striated muscle. Peroxidasin first appears along with collagen IV at the divergence of Placozoa (*Trichoplax adhaerens*) and Choanozoa; however, the requisite Met^{93} and Hyl^{211} residues are absent in *Trichoplax* (Fig. 2). Importantly, the occurrence of the sulfilimine cross-link (*vide supra*) together with peroxidasin throughout Eumetazoa establishes peroxidasin as a primordial oxidant generator embedded within basement membranes that produces hypohalous acids as strong oxidant intermediates for the formation of sulfilimine cross-links (21, 22).

Among eight cnidarians species investigated, *H. magnipapillata* is the one exception that does not contain the sulfilimine bond. This species does not contain Met^{93} and Lys/Hyl^{211} residues or peroxidasin, features that are required for bond formation (Fig. 2), but it does contain peroxinectin, an invertebrate heme peroxidase that is known to be involved in innate immunity (28). These findings indicate that the cnidarian stem ancestor possessed the bond and suggests that *Hydra* underwent a secondary gene loss, a phenomenon previously described for *Hydra* (29). The mesoglea of *Hydra* is complex, composed of both basement membrane and an interstitial matrix of various collagens (30), and it seems designed for flexibility rather than mechanical strength (31). Nondenaturing, nonreducing buffers do not solubilize collagen IV scaffolds in other metazoans but easily extract *Hydra* collagen IV, suggesting a general absence of cross-links (32). Although the relative fragility of the mesoglea is tolerated in *Hydra*, the advent of mesoderm and the emergence of muscular tissue in Bilateria likely required additional mechanical strength, which was provided by the sulfilimine bond, enabling the evolution of complex tissues for locomotion and increased body size.

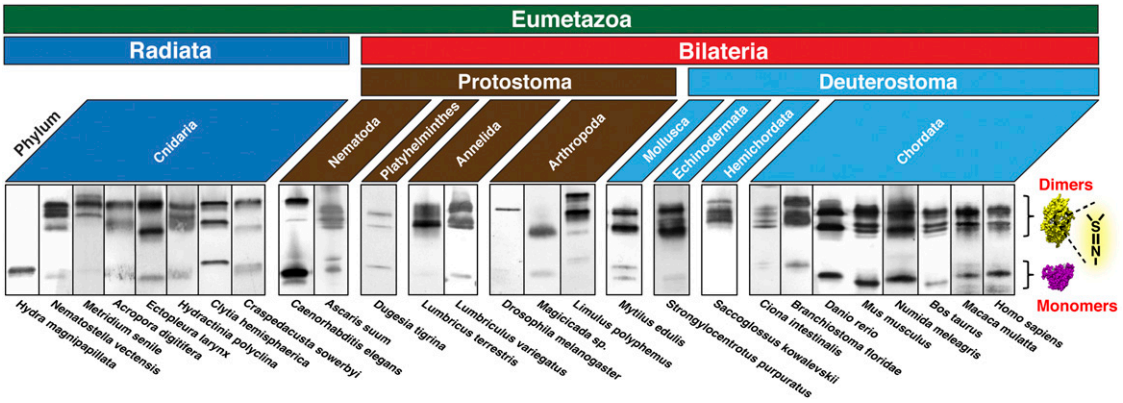

Fig. 3. NC1 hexamers were excised from animal basement membranes and analyzed by SDS/PAGE as shown in Fig. 1 *A* and *B*. The dimeric subunits, which indicate the presence of the bond, were found in nine major eumetazoan phyla. Among eight cnidarians investigated, only *Hydra* NC1 lacked dimeric subunits. All NC1s were immunoblotted against the rat monoclonal antibody, JK2, except for *C. elegans* (rabbit polyclonal; NW-154) and *Drosophila* (mouse monoclonal; 6G7). Black outlines indicate the locations of cropping for blot images.

Essentiality of the Sulfilimine Cross-Link in Tissue Genesis. Sulfilimine cross-link is essential for invertebrate tissue genesis as recently described for the protostomes *D. melanogaster* and *C. elegans* (21, 33). We extended the investigation of essentiality to vertebrate and deuterostome development using the zebrafish model. The expression levels of collagen IV α1 and α2 chains and peroxidasin were measured during early zebrafish development. Real-time quantitative PCR (qPCR) studies showed up-regulation of all three components during gastrulation (Fig. 4*A*). To probe the importance of the bond in early development, morpholino oligomers were designed to inhibit the translation of peroxidasin and microinjected into fertilized one-cell stage embryos. At 24 h postfertilization (hpf), embryos were assessed and found to display a phenotypic ratio of normal development (20/45 embryos), partial curved trunk (21/45 embryos), and severely defective (4/45 embryos), which consisted of cardiac edema, decreased eye size, and gross trunk patterning effects (Fig. 4 *D*–*F*). To determine the effect of the inhibitory morpholinos on bond formation as measured by the presence of cross-linked NC1 dimers, whole embryos were digested with collagenase, and the excised NC1 hexamers were analyzed by SDS/PAGE. The phenotype with partial curved trunk presented with a total absence of dimer bands, indicative of decreased sulfilimine cross-links, as well as a decreased content of collagen IV (Fig. 4*G*). These results indicate that the developmental expression of the cross-linked collagen IV scaffold occurs during gastrulation and that the scaffold is of functional importance for organogenesis.

The developmental timeline in zebrafish for the functional importance of the collagen IV scaffold is congruent with the developmental timelines for several other eumetazoans. Collagen IV KO embryos in *C. elegans* show lethality at the threefold stage and in mouse at embryonic age (days post coitum) 10.5–11.5 (34, 35). For peroxidasin KOs, *C. elegans* embryos arrest at the threefold stage, and *Drosophila* embryos are lethal at the third instar larvae stage (21, 33). In *N. vectensis*, collagen IV and peroxidasin expression are up-regulated during gastrulation (Fig. S9). Collectively, these findings indicate the essentiality of the cross-linked collagen IV scaffold for organogenesis, a stage when mechanical forces manifest in metazoan development.

Conclusions

The triad—a collagen IV scaffold with sulfilimine cross-links, peroxidasin, and hypohalous acids—is a primordial innovation of the basement membrane that is essential for organogenesis and evolution of tissues. The cross-links confer (*i*) stability and tensile strength, enabling the collagen IV scaffold to serve as an anchor for cell attachment and tissue compartmentalization, a ligand for cell surface receptors that modify cell behavior, and a locus for bone morphogenetic protein gradients for patterning in tissue development, and (*ii*) structural integrity to tissues for locomotion and large body size. The cross-linked scaffold, in part, enables the evolution of complex tissues that overcome mechanical forces and the constraint of chemical diffusion in the delivery of nutrients to distant organs in large body size.

Peroxidasin is a primordial oxidant generator embedded within basement membranes that produces hypohalous acids as strong oxidant intermediates for formation of the sulfilimine cross-links. This anabolic function of hypohalous acids is paradoxical, because the canonical role of these acids occurs in vertebrates, where they are produced by myeloperoxidase and eosinophil peroxidase and function as antimicrobial agents in innate immunity (21, 22, 27, 28). Furthermore, peroxidasin is the only animal heme peroxidase spanning both invertebrates and vertebrates (28), and it shares a peroxidasin-like ancestor with the vertebrate peroxidases: myeloperoxidase, eosinophil peroxidase, and lactoperoxidase (27). Thus, in invertebrates, the hypohalous acids formed by peroxidasin may have a dual function of forming sulfilimine cross-links and serving as a primitive form of innate immunity. Indeed, mosquito gut peroxidasin is up-regulated after bacterial infection, and its knockdown reduces bacterial clearance and host survival (36).

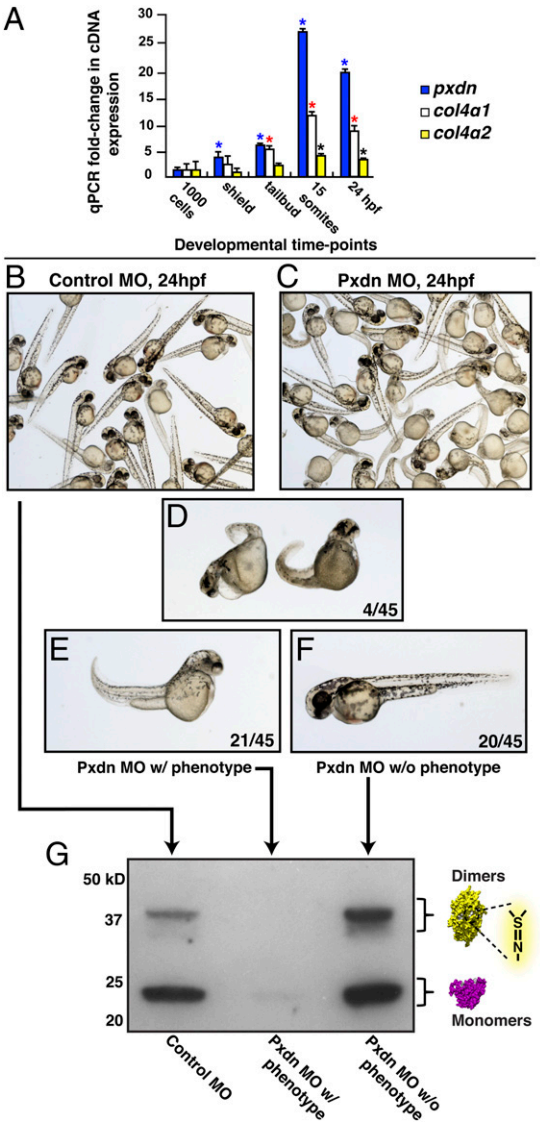

Fig. 4. Expression of collagen IV and Pxdn during development in zebrafish and morpholino (MO) knockdown of peroxidasin in zebrafish embryos. (*A*) *Pxdn* and *collagen IV* expression during zebrafish embryonic development. Real-time qPCR studies were conducted to examine expression levels of *Pxdn*, *collagen4α1*, and *collagen4α2*. *Student *t* test *P* value < 0.03 compared with expression at 1,000 cells. Error bars = SEM. Blue, pxdn; red, col4a; black, col4a2. (*B*) Control and (*C*) Pxdn MO groups. MO-injected embryos displayed (*D*) general severe defects that include cardiac edema, smaller eyes, and gross trunk patterning defects (4/45), (*E*) partial curved trunk (21/45), or (*F*) normal development (20/45). (*G*) SDS/PAGE analysis of Pxdn MO embryonic phenotypes at 24 hpf by Western blot. Collagenase digests were normalized for total protein load by protein stain with SYPRO-Ruby (Fig. S8).

Materials and Methods

Animals. *H. magnipapillata* cultures of strain 105 were cultured in the laboratory by R.E.S. *Acropora digitifera* was purchased from Happy Coral, Inc. *Hydractinia polyclina* was collected from Clark Cove in Mount Desert, ME. *Ectopleura larynx* and *Metridium sp.* were collected from the piers at the Town Dock in Northeast Harbor, ME. *C. elegans* was cultured in our laboratory. *L. terrestris* was purchased from NASCO. *Lumbriculus variegatus* was purchased from Aquatic Foods. *D. melanogaster* was obtained from John and Lisa Fessler (University of California, Los Angeles, CA). *Dugesia tigrina* was purchased from Ward's Natural Science. *Mytilus edulis* was purchased from a local seafood market in Nashville, TN. *Strongylocentrotus purpuratus* was purchased from M-REP. *Saccoglossus kowalevskii* was purchased from Marine Biological Laboratory. *Ciona intestinalis* and *Branchiostoma floridae*

were purchased from Gulf Specimen Marine Laboratory. *D. rerio* was purchased from Aquatic Critter and cultured in the laboratory by V.P.Y. *Mus musculus* kidneys were purchased from Pel-Freez Biologicals. *M. mulatta* kidney tissue was donated by Heather Hudson (Kansas University Medical Center, Kansas City, KS). Human kidney tissues were obtained from normal donors from the National Disease Research Interchange. Bacterial collagenase was purchased from Worthington Biochemical. JK2 rat monoclonal antibody was purchased from Yoshikazu Sado (Shigei Medical Research Institute, Okayama, Japan). *C. elegans* anticollagen IV rabbit polyclonal antibody, NW-154, was donated by James Kramer (Northwestern University, Evanston, IL). *Drosophila* anticollagen IV mouse monoclonal antibody, 6G7, was donated by John and Lisa Fessler.

Isolation and Purification of Collagen IV NC1 Hexamers. Two primary methods were used for isolation and purification of collagen IV NC1 hexamer because of variance in Bilaterian and Cnidarian tissue. Bilaterian (Deuterostome and Protostome) tissues were frozen in liquid nitrogen, pulverized in a mortar and pestle with liquid nitrogen, and then, sonicated in 1% (wt/vol) deoxycholate plus 10 mM EDTA (pH 8.0) and 10 mM Tris·Cl (pH 8.0); the insoluble material was isolated after centrifugation at 10,000 × *g* for 15 min. The pellet was then extracted with 1 M NaCl plus 50 mM Tris·Cl (pH 7.5), and the insoluble material was isolated after centrifugation at 10,000 × *g* for 15 min. The pellet was then extracted a final time with cold ddH20, and the insoluble material was isolated after centrifugation at 10,000 × *g* for 15 min. The pellet was dispersed into 2.0 mL g^{-1} buffer [50 mM Hepes (pH 7.5), 10 mM $CaCl_2$, 1 mM PMSF, 5 mM benzamadine-HCl, 25 mM 6-amino-n-hexanoic acid] containing 0.1 mg mL^{-1} bacterial collagenase (Worthington Biochemical). The mixture was incubated at 37 °C with stirring for 24 h. Collagenase-solubilized material was dialyzed against 50 mM Tris·Cl (pH 7.5). Cnidarian tissues were frozen in liquid nitrogen, pulverized in a mortar and pestle, and then, homogenized in 2.0 mL g^{-1} digestion buffer and 0.1 mg mL^{-1} bacterial collagenase; they were allowed to digest at 37 °C with spinning for 24 h. Liquid chromatography purification of solubilized NC1 varied by species based on protein yield. All excised deuterostome and protostome NC1s were purified by anion exchange chromatography (DE-52 cellulose or GE HiTrap Q HP) and gel exclusion chromatography (GE Superdex 200 10/300 GL).

SDS/PAGE and Immunoblot Analysis of NC1 Hexamers. Collagenase-solubilized NC1 hexamers were analyzed by SDS/PAGE in 12% (wt/vol) *bis*-acrylamide minicells with Tris-Glycine-SDS running buffer. Sample buffer was 62.5 mM Tris·HCl (pH 6.8), 2% (wt/vol) SDS, 25% (wt/vol) glycerol, and 0.01% bromophenol blue. Western blotting of SDS-dissociated NC1 hexamer was done with the rat mAb JK-2, except for *Drosophila* and *C. elegans* NC1 hexamers, which were assayed by 6G7 (mouse monoclonal; *Drosophila* anticollagen IV antibody) and NW-154 (rabbit polyclonal; *C. elegans* anticollagen IV antibody), respectively. All Western blotting in Fig. 3 was done with Alkaline Phosphatase substrate, and zebrafish mutant Western blotting in Fig. 4 was done with Thermo-Scientific SuperSignal West Femto chemiluminescent substrate and developed on film. Protein staining was done with Coomassie Brilliant Blue (G-250) or Invitrogen SYPRO-Ruby.

Next Generation RNA Sequencing. Transcriptomes used in this study were sequenced at the Vanderbilt Technologies for Advanced Genomics Core Facility using a custom pipeline (full method is in *SI Materials and Methods*). De novo assembly of transcriptomes was performed using Velvet/Oases, Trinity and CLC Genomic Workbench software packages with default settings. The accuracy of de novo assembly was checked in a parallel next generation RNA-Seq experiment using RNA from mouse PFHR9 cells. De novo assembled transcripts were used to generate BLAST databases to search for collagen IV similar sequences using *tblastn* with e-value cutoff set to 10^{-15}. Multiple sequence alignments were performed with the Geneious v5-6 software (Biomatters) built-in algorithm. To identify peroxidasin (pxdn) sequences for corresponding species, we followed the same procedure using mouse pxdn as a query. Candidates for collagen IV and pxdn were then checked with the NCBI's Conserved Domain Database to ensure that correct sequence information was used in the study.

Isolation of Sulfilimine Cross-Linked Peptides. NC1 hexamers from various animals were denatured by boiling in 0.2 M Tris·HCl buffer (pH 7.5) containing 4 M Guanidine-HCl and 25 mM DTT. The proteins were alkylated with iodoacetamide, precipitated with ethanol, resuspended in freshly made ammonium bicarbonate, and digested with sequencing-grade modified trypsin (Promega). The tryptic digest was fractionated on a Superdex peptide column (Amersham Biosciences), and the fractions containing polypeptides of 3,000–6,000 molecular weight (~9 mL column volume) were pooled together, freeze-dried, and stored until MS analyses were performed.

High-Resolution FTMS of Sulfilimine Cross-Linked Peptides. Dry samples were reconstituted in 0.1% formic acid and loaded onto a capillary reverse phase analytical column (360 μm o.d. × 100 μm i.d.) using an Eksigent NanoLC Ultra HPLC and autosampler. The analytical column was packed with 20 cm C18 reverse phase material (Jupiter; 3-μm beads, 300 Å; Phenomenex) directly into a laser-pulled emitter tip. Peptides were gradient-eluted at a flow rate of 500 nL/min, and the mobile phase solvents consisted of 0.1% formic acid and 99.9% water (solvent A) and 0.1% formic acid and 99.9% acetonitrile (solvent B). A 90-min gradient was performed and consisted of 0–10 min (sample loading), 2% B; 10–50 min, 2–35% B; 50–60 min, 35–90% B; 60–65 min, 90% B; 65–70 min, 2–90% B; 70–90 min (column equilibration), 2% B. On gradient elution, peptides were mass analyzed on either an LTQ Orbitrap XL or LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific); each mass spectrometer is equipped with a nanoelectrospray ionization source. The instruments were operated either using a data-dependent method or with a targeted method to enable analysis of low-abundant cross-linked NC1 peptides. For data-dependent analyses, full scan (*m/z* = 400–2,000) spectra were acquired with the Orbitrap as the mass analyzer (resolution = 60,000), and the three most abundant ions in each MS scan were selected for fragmentation in the LTQ, which was followed by data-dependent MS^3 analysis of the three most abundant ions in each MS^2 scan. An isolation width of 2 *m/z*, an activation time of 30 ms, and 35% normalized collision energy were used to generate MS^2 and MS^3 spectra. The MS^n AGC target value was set to $1e^4$ or $2e^4$, and the maximum injection time was either 150 or 250 ms, respectively. For MS^2 scans, a minimum threshold of 500 was used to trigger data-dependent spectra, whereas a threshold of 500 or 100 was used to trigger MS^3 spectra. Dynamic exclusion was enabled, with an exclusion duration of either 20 (LTQ Orbitrap Velos) or 60 s (LTQ Orbitrap XL). For analyses where cross-linked peptides were too low in intensity to be selected data-dependently, targeted methods were performed. Full scan spectra were similarly acquired in the Orbitrap, but specific *m/z* values were provided in the data acquisition method to facilitate MS^2 and MS^3 fragmentation irrespective of the intensity of precursors.

FTMS Peptide Identification and Bioinformatics. Two strategies were followed for the identification of sulfilimine cross-linked peptides. First, raw data files were manually searched for sulfilimine cross-linked peptides using the Xcalibur 2.2 QualBrowser software (Thermo Scientific). To calculate the monoisotopic mass of the sulfilimine cross-linked peptides, the mass of two hydrogen atoms was subtracted from the sum of the masses for Met^{93}- and Lys^{211}-containing peptides, each generated using GPMAW version 8.00sr1 (Lighthouse Data) as previously shown in the work by Vanacore et al. (20). Second, sulfilimine cross-linked peptides not found manually were identified by searching the bovine, mouse, *D. rerio*, *D. melanogaster*, *C. elegans*, or *N. vectensis* subsets of the UniprotKB database (www.uniprot.org) using either SEQUEST (Thermo Scientific) on a high-speed, multiprocessor Linux cluster in the Advanced Computing Center for Research and Education at Vanderbilt University or the Myrimatch algorithm (37) on a two six-core Intel Xeon processor HP Proliant DL160 G6 server running Windows server 2008 R2 enterprise using the Bumbershoot suite (38). The search files were adapted for the identification of peptides that included the amino acid modifications of interest [i.e., −48.0034 on Met and +45.9877 or +61.9826 on Lys as well as carbamidomethylation of cysteine (+57.0214) and oxidation of methionine (+15.9949)]. The result files obtained from SEQUEST and/or Myrimatch were assembled using IDPicker (39) and/or Scaffold 3.6.4 (Proteome Software). MS^3 spectra matching peptides containing modified Lys^{211} or Met^{93} were confirmed by either manual evaluation or processing the raw data files through ScanRanker (40), and the spectra were annotated with IonMatcher.

qPCR Analysis of Embryonic Expression of Zebrafish Pxdn and Collagen IV. WT Ekkwill embryos were collected, raised in E2 medium at 28 °C, and staged to the appropriate developmental phase. At 1,000 cells, shield, tailbud, 15 somites, and 24 hpf, at least 50 embryos were harvested for total RNA isolation using Tri-Reagent (Sigma-Aldrich). Duplicate cDNA synthesis was performed with 250 ng total RNA using the Supermix cDNA synthesis kit (Quanta Biosciences) in accordance with the manufacturer's protocol. For real-time qPCR analyses, triplicate reactions were used for each cDNA synthesis, and crossing thresholds were determined with SYBR green methods (Roche) (41). Expression levels of Pxdn and collagen IVα1 and -α2 were normalized to β-actin levels at each developmental stage. Fold change in gene expression during development was determined relative to the 1,000-cell state. qPCR oligo pairs used are listed in *SI Materials and Methods*.

Morpholino Knockdown of Pxdn in Zebrafish Development. Antisense morpholino directed against Pxdn (5′-AGT TTC GCA CAG TCC GCA ACG CCA T-3′) and standard control (5′-CCT CTT ACC TCA GTT ACA ATT TAT A-3′) morpholino were resuspended in nuclease-free water, and ∼1 nL 1 mM solution was microinjected in the cytoplasm of WT Ekkwill one-cell embryos. Microinjected embryos were raised in E2 medium for ∼24 h at 28 °C before manual dechorionation with forceps for phenotypic characterizations. Embryos were imaged in 2% methylcellulose using an Olympus MXV110 stereomicroscope. For Western blots, control and Pxdn morpholino-microinjected embryos were scored at 24 hpf and snap-frozen in liquid nitrogen.

Aspirnauts (Middle School, High School, and Undergraduate Students). The students were sponsored by the Aspirnaut K-20 Science, Technology, Engineering, and Math (STEM) pipeline (www.aspirnaut.org) for diversity, which focuses on elevating STEM achievement of students from geographically and economically disadvantaged backgrounds, students from underrepresented racial and ethnic minority groups, and Native Americans. In this study over five summer periods, students participated as teams in a search for the sulfilimine bond. Students were educated on the topic, including basic biology and chemistry, and trained in basic biochemical technologies sufficient for the investigation. They isolated, purified, and analyzed collagen IV NC1 hexamers from 27 animal species. In addition, they participated individually in separate research projects. These activities, conducted as a member of an ongoing research team, provided an in-depth research experience, involving crafting of questions, designing and conducting experiments, and presenting and arguing results while contributing to the advancement of a research topic. At the end of the summer internship, students prepared written reports and gave oral and poster presentations to their peers and mentors. The names and demographics of the students are presented in Table S2.

ACKNOWLEDGMENTS. The technical work of Parvin Todd, Mohamed Rafi, Neonila Danylevych, Salisha Hill, and Ashley Smith is greatly appreciated. We acknowledge the Vanderbilt University Mass Spectrometry Research Center and Proteomics Laboratory for use of their instruments. We thank Michelle Bailey for assistance with numerous animal species, Dr. Evelyn Houliston and Tsuyoshi Momose for *Clytia hemisphaerica* cultures as well as access to *Clytia* transcriptome data, Dr. Allison Smith for field collection of *Craspedacusta sowerbyi*, Julijana Ivanisevic for field collection of *Oscarella sp.*, and Drs. Paulyn Cartwright, Carol Vines, Ann Tarrant, Adam Reitze, and Athula Wikramanayake for *Nematostella vectensis*. This work was supported by National Institute of Diabetes and Digestive and Kidney Diseases Short-Term Research Experience for Underrepresented Persons, Freytag Holdings LLC, Vanderbilt University Medical Center, Vanderbilt Center for Matrix Biology, Vanderbilt–Ingram Cancer Center, Vanderbilt Division of Nephrology Faculty Development Fund (to R.M.V.), Mount Desert Island Biological Laboratory Salisbury Cove Research Fund and an F. H. Epstein Fellowship (to B.G.H.), contributions by Hector High and Knowledge Is Power Program Delta schools, and gifts to The Aspirnaut Program. This work was also supported by National Institutes of Health Grants R25 DK09699-02, R01 DK18381 (to B.G.H.), R15 DK091009-01 (to M.T.I.), and American Recovery and Reinvestment Act Supplement Grant 2P01 DK065123-07 (to B.G.H.).

1. Petersen OW, Rønnov-Jessen L, Howlett AR, Bissell MJ (1992) Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. *Proc Natl Acad Sci USA* 89(19): 9064–9068.
2. Lukashev ME, Werb Z (1998) ECM signalling: Orchestrating cell behaviour and misbehaviour. *Trends Cell Biol* 8(11):437–441.
3. Daley WP, Yamada KM (2013) ECM-modulated cellular dynamics as a driving force for tissue morphogenesis. *Curr Opin Genet Dev* 23(4):408–414.
4. Hynes RO (2009) The extracellular matrix: Not just pretty fibrils. *Science* 326(5957): 1216–1219.
5. Hynes RO (2012) The evolution of metazoan extracellular matrix. *J Cell Biol* 196(6): 671–679.
6. Yurchenco PD (2011) Basement membranes: Cell scaffoldings and signaling platforms. *Cold Spring Harb Perspect Biol* 3(2):pii:a004911.
7. Ozbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker RP, Adams JC (2010) The evolution of extracellular matrix. *Mol Biol Cell* 21(24):4300–4305.
8. Bryant DM, Mostov KE (2008) From cells to organs: Building polarized tissue. *Nat Rev Mol Cell Biol* 9(11):887–901.
9. Vracko R (1974) Basal lamina scaffold-anatomy and significance for maintenance of orderly tissue structure. *Am J Pathol* 77(2):314–346.
10. Orlando G, et al. (2013) Discarded human kidneys as a source of ECM scaffold for kidney regeneration technologies. *Biomaterials* 34(24):5915–5925.
11. Song JJ, Ott HC (2011) Organ engineering based on decellularized matrix scaffolds. *Trends Mol Med* 17(8):424–432.
12. Maher B (2013) Tissue engineering: How to build a heart. *Nature* 499(7456):20–22.
13. Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG (2003) Alport's syndrome, Goodpasture's syndrome, and type IV collagen. *N Engl J Med* 348(25):2543–2556.
14. Gould DB, et al. (2005) Mutations in Col4a1 cause perinatal cerebral hemorrhage and porencephaly. *Science* 308(5725):1167–1171.
15. Gould DB, et al. (2006) Role of COL4A1 in small-vessel disease and hemorrhagic stroke. *N Engl J Med* 354(14):1489–1496.
16. Pedchenko V, et al. (2010) Molecular architecture of the Goodpasture autoantigen in anti-GBM nephritis. *N Engl J Med* 363(4):343–354.
17. Moser M, Legate KR, Zent R, Fässler R (2009) The tail of integrins, talin, and kindlins. *Science* 324(5929):895–899.
18. Wang X, Harris RE, Bayston LJ, Ashe HL (2008) Type IV collagens regulate BMP signalling in Drosophila. *Nature* 455(7209):72–77.
19. Sawala A, Sutcliffe C, Ashe HL (2012) Multistep molecular mechanism for bone morphogenetic protein extracellular transport in the Drosophila embryo. *Proc Natl Acad Sci USA* 109(28):11222–11227.
20. Vanacore R, et al. (2009) A sulfilimine bond identified in collagen IV. *Science* 325(5945):1230–1234.
21. Bhave G, et al. (2012) Peroxidasin forms sulfilimine chemical bonds using hypohalous acids in tissue genesis. *Nat Chem Biol* 8(9):784–790.
22. Weiss SJ (2012) Peroxidasin: Tying the collagen-sulfilimine knot. *Nat Chem Biol* 8(9): 740–741.
23. Hung CH, Butkowski RJ, Hudson BG (1980) Intestinal basement membrane of Ascaris suum. Properties of the collagenous domain. *J Biol Chem* 255(10):4964–4971.
24. Lunstrum GP, et al. (1988) Drosophila basement membrane procollagen IV. I. Protein characterization and distribution. *J Biol Chem* 263(34):18318–18327.
25. Guo XD, Kramer JM (1989) The two Caenorhabditis elegans basement membrane (type IV) collagen genes are located on separate chromosomes. *J Biol Chem* 264(29): 17574–17582.
26. Sarras MP, Jr., et al. (1991) Extracellular matrix (mesoglea) of Hydra vulgaris. I. Isolation and characterization. *Dev Biol* 148(2):481–494.
27. Soudi M, Zamocky M, Jakopitsch C, Furtmüller PG, Obinger C (2012) Molecular evolution, structure, and function of peroxidasins. *Chem Biodivers* 9(9):1776–1793.
28. Zamocky M, Jakopitsch C, Furtmüller PG, Dunand C, Obinger C (2008) The peroxidase-cyclooxygenase superfamily: Reconstructed evolution of critical enzymes of the innate immune system. *Proteins* 72(2):589–605.
29. Steele RE, David CN, Technau U (2011) A genomic view of 500 million years of cnidarian evolution. *Trends Genet* 27(1):7–13.
30. Shimizu H, et al. (2008) The extracellular matrix of hydra is a porous sheet and contains type IV collagen. *Zoology (Jena)* 111(5):410–418.
31. Sarras MP, Jr., Deutzmann R (2001) Hydra and Niccolo Paganini (1782–1840)—two peas in a pod? The molecular basis of extracellular matrix structure in the invertebrate, Hydra. *BioEssays* 23(8):716–724.
32. Fowler SJ, et al. (2000) Characterization of hydra type IV collagen. Type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose. *J Biol Chem* 275(50):39589–39599.
33. Gotenstein JR, et al. (2010) The C. elegans peroxidasin PXN-2 is essential for embryonic morphogenesis and inhibits adult axon regeneration. *Development* 137(21): 3603–3613.
34. Gupta MC, Graham PL, Kramer JM (1997) Characterization of alpha1(IV) collagen mutations in Caenorhabditis elegans and the effects of alpha1 and alpha2(IV) mutations on type IV collagen distribution. *J Cell Biol* 137(5):1185–1196.
35. Pöschl E, et al. (2004) Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. *Development* 131(7):1619–1628.
36. Garver LS, Xi Z, Dimopoulos G (2008) Immunoglobulin superfamily members play an important role in the mosquito immune system. *Dev Comp Immunol* 32(5):519–531.
37. Tabb DL, Fernando CG, Chambers MC (2007) MyriMatch: Highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res* 6(2):654–661.
38. Holman JD, Ma ZQ, Tabb DL (2012) Identifying proteomic LC-MS/MS data sets with Bumbershoot and IDPicker. *Curr Protoc Bioinformatics* 13(2012): 13.17.
39. Ma ZQ, et al. (2009) IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J Proteome Res* 8(8):3872–3881.
40. Ma ZQ, et al. (2011) ScanRanker: Quality assessment of tandem mass spectra via sequence tagging. *J Proteome Res* 10(7):2896–2904.
41. Yin VP, et al. (2008) Fgf-dependent depletion of microRNA-133 promotes appendage regeneration in zebrafish. *Genes Dev* 22(6):728–733.